

Designing Scientific Analysis Tools for Collaborative Environments

Velvin R. Watson

NASA Ames Research Center

vwatson@arc.nasa.gov

It will be assumed that this audience concurs that highly interactive visualization of high resolution dynamic representations of data is critical for some scientific analyses, and that including this capability within the collaborative environment for scientists is as important as including video, audio, whiteboards, chat rooms, and document sharing. Therefore, this presentation will be devoted to design decisions necessary to achieve this capability in the scientists' collaborative environments. Some of these decisions are counter to the currently popular decisions, and they should inspire a good discussion during the workshop.

Design decisions to achieve acceptable latency

The need for some peer to peer communications between sites

A popular trend now for collaboration environments is to base them on Web technology. It is convenient to start collaborations with browsers on the Web, but there are instances where one needs to leave Web technology and use some peer to peer communications to achieve acceptable latency within a collaborative session between scientists.

One such instance is for synchronous collaboration with highly interactive dynamic analyses. The displays for all sites must be nearly synchronous so that the dynamical events being displayed can be discussed as they occur. It is unlikely that Web-based architectures that pass all communications through HTTP servers will be able to pass events between sites with sufficiently small delays to provide suitably synchronized displays. On the other hand, tests show that a peer to peer architecture that provides dedicated event handling ports between the remote sites serviced by continuously running daemons does provide reasonably synchronized displays.

The need for local displays to be driven by local controls

Latency between user controls and displays is even more critical. Dynamically "driving through" scenes of data and tracking a dynamic physical phenomena, similar to driving a simulated race car in a computer game, requires a very close coupling between the display and the controls. Therefore, for highly interactive analyses, local displays should be driven directly by local controls rather than driven via round trip communications with a remote site. Collaborative environment architectures requiring remote round trip communications between controls and displays are not likely to offer sufficiently rapid control of the display for tracking complex dynamic phenomena.

Design decisions to achieve display of high resolution dynamic scenes of the data

The need for local rendering and intelligent, compact communications between remote sites

Another popular design decision is to use the "infinite bandwidth" approach and have rendering done on a remote server. However, collaborative environment architectures that utilize rendering on a remote server and then send pixels to the scientists' local computers cannot currently provide the dynamic scenes that are required for some types of scientific analyses with the network bandwidths that are commonly available to scientists.

There will be situations where sending pixels from remote servers may be the best approach, but the designer should be aware that selecting this approach eliminates the possibility of providing high resolution, dynamic graphics with the network bandwidths that are commonly available between scientists. The basic problem is that the bandwidth of visual information that can be displayed from current computer monitors to the users' eyes (and that is needed for high resolution dynamic scenes) is more than an order of magnitude greater than the typical network bandwidth between scientists. Furthermore, we are not even close to exceeding the information processing bandwidth of the human visual system, and future systems will increase the bandwidth of visual information to the users' eyes. The human eye has one hundred times more receptors than pixels on the current workstation displays[30]. Therefore, even if we knew how to efficiently map pixel information to receptors, we could increase the bandwidth between the computer and the user by one hundred without exceeding the information processing bandwidth of the human visual system.

Network bandwidths are increasing rapidly, and some argue that networks will soon provide the bandwidths to send pixels over the network fast enough to equal the information bandwidth between the workstation and the user. However, this is not likely to occur very soon because the number of users sharing this increased bandwidth will increase and the information bandwidth between the workstation and each user will also increase. Designers are using the ever increasing workstation processing power and display technology to provide larger and more sophisticated displays (such as autostereoscopic displays) and other types of sensory information.

Some would also argue that compression algorithms can now be used to reduce the network bandwidth required by orders of magnitude. For example, transmission of MPEG movies appears to give good quality movies. However, unlike movie viewers, scientists are reluctant to compress scenes of scientific data by more than an order of magnitude for fear of creating artifacts that may be misinterpreted as a scientific phenomena.

A current trend is to free scientists from their tethers by providing wireless networking. This shift to wireless networking may mean that we should be designing for lower bandwidths instead of greater bandwidths for the near future. (Actually, the broad spectrum of bandwidths that scientists will use suggests that we should design systems that are network aware and change to accommodate a wide range of negotiated bandwidths.)

Collaborative environment architectures that send window drawing commands (such as the ITU T.120 architecture) also do not provide adequate performance for remote applications that require dynamic, high resolution graphics. To illustrate this, the reader can try to use Microsoft NetMeeting to share any application that uses dynamic, high resolution graphics.

Collaborative environment architectures that use local rendering and send intelligent, compact communications between remote sites (for example application specific data and events) can provide the high resolution dynamic scenes required for scientific visual analysis if the client computers at each site can render the scenes fast enough. (Fortunately, even the PCs are gaining the ability to render high resolution 3D scenes rapidly.)

This local rendering approach for providing high performance graphics in a collaborative mode requires distribution of the data to be analyzed to the remote sites. For large data sets with low Internet bandwidth, the time to transfer the data may be large. In this case, the data sets should be distributed to the remote sites before the collaborative analysis session so that the data transfer does not impact the interactivity of the analysis. (Note, that this would eliminate the capability to continuously steer a simulation or experiment.)

The need for paging of data extracts for analyzing large data sets

For extremely large data sets, distributing all of the data to the remote sites may not be feasible. An approach for this case could be to transfer only extracts of the data and use the highly interactive analysis on each extract. For example, one could first transfer and analyze the data over the whole domain but only on a coarse grid. Then data on a finer grid in regions of specific interest could be transferred and analyzed next.

The need for rich user interfaces and powerful local computing at each remote site

Another popular design decision is to design the collaboration environment for "lowest common denominator" clients and computers at the scientists' sites. However, for highly interactive analysis of complex dynamic data, a rich, user interface on a computer with sufficient power to render dynamic high resolution 3D scenes is much more effective. In the future, computers will be capable of providing user interfaces with voice recognition, user tracking, user awareness, support for haptic devices, and other features. For scientific research, it is still highly cost effective to invest in improving the capabilities of the local computer and the computer-user interface. Designing collaborative environments for low power "lowest common

denominator” clients and computers will prevent maximizing the potential for effective analysis through the use of advanced interfaces.

Design decisions to achieve the capability to publish and experience analysis sessions for asynchronous collaboration

The need for a tailored scripting language and journal files

The capability for authors to publish explorations through data (recorded analysis sessions that can be played back) and for remote colleagues to experience the explorations can be easily achieved if the scientific analysis tool has a tailored scripting language and a journal file capability. This has been demonstrated as described below.

Demonstration of collaborative scientific analysis tools based on these design decisions

The basis for the tools

An asynchronous collaboration tool, FASTexpeditions[22], and a synchronous collaboration tool, RemoteFAST[22], were created based on FAST[21] (Flow Analysis Software Toolkit), which is a stand alone scientific visualization tool.

The tailored scripting language and journal files

Within FAST, all user actions have an associated script command that is automatically recorded during an analysis session. These recorded script commands (the journal file) can then be used to drive FAST so that the analysis can be replayed. These journal files are ASCII text files which can be easily edited with any text editor.

Publishing and experiencing analysis sessions on the Web for asynchronous collaboration

The asynchronous collaboration tool, FASTexpeditions, was created using this journal file capability of FAST. The scientific data and various journal files containing expeditions through the data are published as FASTexpeditions within a scientist’s Web report. When one of these FASTexpeditions is selected by a viewer, the data is automatically loaded onto the viewer’s local computer, FAST is automatically started on the local computer, and a journal file to set up the initial scene is automatically executed. Subsequent selections of the expeditions from the Web report causes execution of the journal files for those expeditions.

For most of the investigations that we have posted to the Web, all of the expeditions (journal files) are packaged and downloaded along with the initial data because doing this permits playing of any expedition without returning to the remote Web server for the journal files. In this case, the URL used on the Web page refers to the downloaded journal files on the local disk, so the Web browser gets these immediately from the local computer disk rather than waiting for the remote Web server to respond and deliver them.

Sound files can be included in the journal files for an audio description of the analysis as it occurs.

The use of peer to peer communications between sites for synchronous collaboration

The synchronous collaboration tool, RemoteFAST, was created using dedicated event handling ports serviced by continuously running daemons listening for events passed between sites.

The use of local rendering and intelligent, compact communications between remote sites

During a collaborative session, the data is first transferred to the local computer and FAST is started (usually using the FASTexpeditions described above). Then the peer to peer communications is initiated (usually by selecting the “RemoteFAST” selection on a Web page). From then on, the script commands from FAST on the controlling site are

automatically sent to all controlled sites to drive FAST at the controlled sites. All scenes are rendered on the local computer by FAST. The script commands are an intelligent, compact form for communication of events between the sites.

The use of a rich user interface and powerful local computing at each remote site

A rich user interface tailored for scientific analysis of fluid dynamics was employed on SGI workstations which render the high resolution dynamic scenes locally. (Note that PCs are now also capable of rendering dynamic high resolution scenes of data.)

Performance of these collaborative scientific analysis tools

RemoteFAST and FASTexpeditions have been tested in collaborative sessions between sites within the U.S. and between sites in different continents. Within the U.S., the tests were conducted primarily between the NASA Ames Research Center in California and the EPA (Environmental Protection Agency) in North Carolina. Tests between the U.S. and Australia were conducted between NASA Ames Research Center and Perth Australia. Tests between the U.S. and Europe were conducted between the EPA and Monte Carlo, Monaco or Poitiers, France. Figure 1 shows the computer screen during a session.

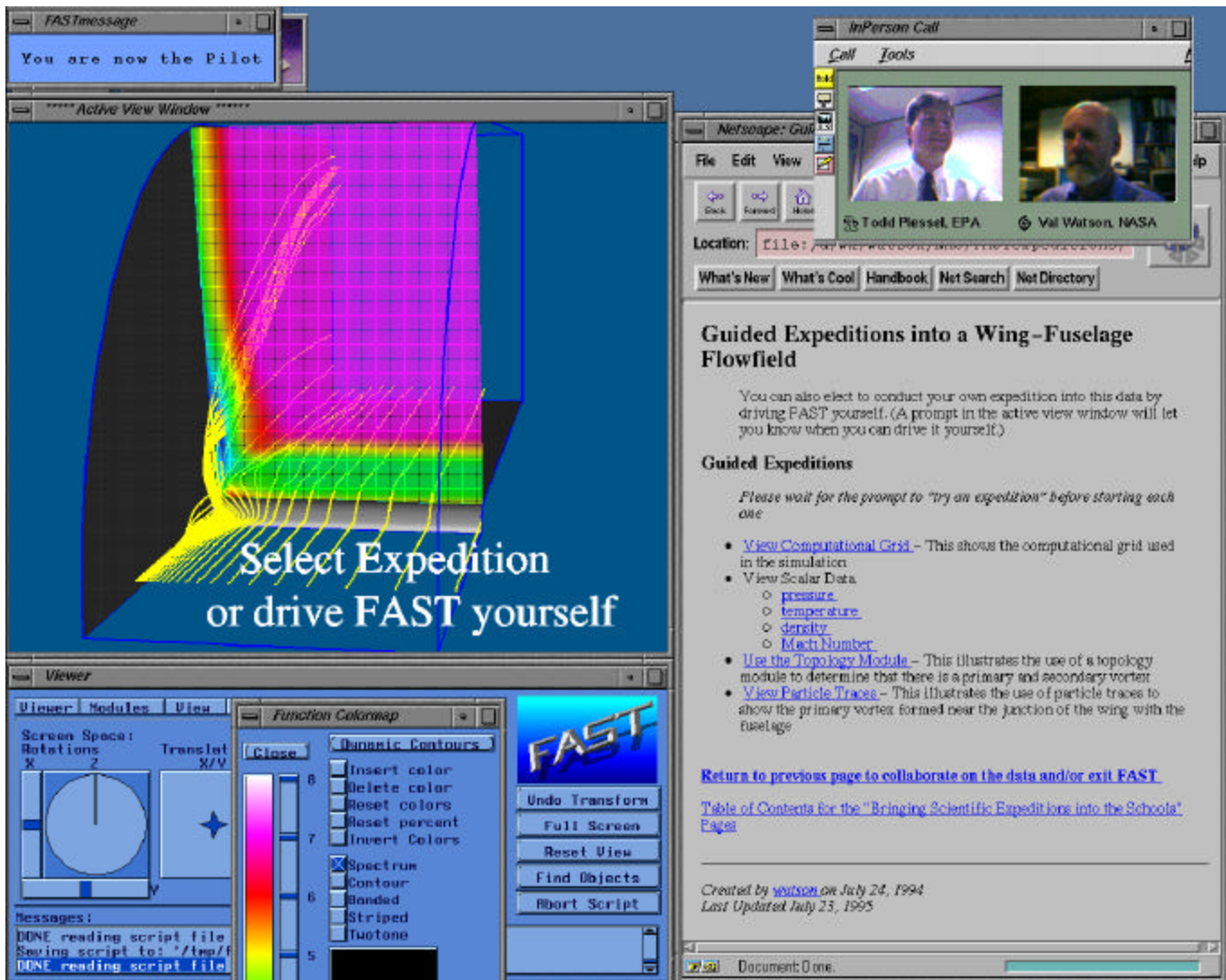


Figure 1. Illustration of a FASTexpedition and RemoteFAST session

RemoteFAST and FASTexpeditions were highly effective for both synchronous and asynchronous collaboration. The effectiveness of the collaboration was nearly as good as being together in the same office and looking at the same workstation while using FAST for the analysis or for a playback of an analysis.

For synchronous collaboration, the response of the visual analysis tool was nearly the same as in stand-alone mode. All sites were able to view the same high resolution (1280x1024), dynamic, 3D scenes simultaneously. Individual sites could independently control their own scene viewing position, but the viewing position could also be resynchronized with the controlling site's viewing position. Control of the analysis was easily transferred between sites. The bandwidth utilized between sites during an interactive collaboration session (after the data was downloaded) was measured to peak at less than 1K bit/second. Note that this low bandwidth utilization and high display performance is achieved by sending script commands over the network and by having the local computer create and render the scenes. This performance cannot now be achieved by sending pixels over the network. Even systems that send scene graphs (such as VRML files) over the network do not match this performance.

For asynchronous collaboration, the analyses posted on the Web were easily downloaded and played. After the initial data download, the playback performance was identical to the performance of playback from journal files on the local workstation disk.

Stereo glasses were often used to obtain stereoscopic scenes in both synchronous and asynchronous modes.

The major advantages of FASTexpeditions over VRML or movie files posted on the Web are:

1. The 3D display performance is superior.
2. Viewers download the actual data and can perform their own “what if” analysis on the data.
3. Viewers can modify the analyses they download and post their own analyses back on the Web.
4. Viewers can collaboratively review and modify the posted analyses with remote colleagues, and these analyses can be posted back onto the Web.

RemoteFAST and FASTexpeditions were used in conjunction with InPersonTM, SGI’s desktop video conference tool, whenever the network bandwidth was high enough (i.e., between France and the U.S. and between sites within the U.S.). Ordinary phones were used instead of InPersonTM when the network bandwidths would not support satisfactory desktop video (i.e., between Monaco and the U.S. and between Australia and the U.S.).

The scenario used most often to demonstrate the features of FASTexpeditions and RemoteFAST follows:

1. A scientist goes to a Web site where FASTexpeditions of various analyses of computer simulations of physics are posted.
2. The scientist selects one of the FASTexpeditions which automatically loads the data and creates an initial scene.
3. The scientist views several of the posted analyses of the data.
4. The scientist then extends one of the author’s posted analysis with his/her own “what if” analysis.
5. The scientist then contacts the author of the posted analyses with a phone or InPersonTM and asks the author about one of the features seen in an analysis.
6. The author and the scientist then both initiate a remote collaboration by making selections on the Web page to automatically start RemoteFAST.
7. The author and the scientist then use RemoteFAST collaboratively to investigate the feature.

Typically, the desktop video was only used at the beginning of the collaborative session when establishing initial contact. When the interest shifted from the initial “hello” to the analysis of the data, the primary focus was shifted to the 3D scenes of the visual analysis process and to the audio.

FAST, is available for download at <http://www.nas.nasa.gov/Software/FAST/>. FASTexpeditions and RemoteFAST are available for download at <http://www.nas.nasa.gov/Software/FAST/FASTexpeditions/>.

References

Pointers to Current Research on Collaborative Environments

Current Research by Organization

Department of Energy

[1] Common Component Architecture (CCA)

(CCA provides components especially valuable for scientific research)

Contact - Rob Armstrong - Sandia

Website - <http://z.ca.sandia.gov/~cca-forum/>

[2] Collaboration Technologies Group

(General research and development in collaboration technologies)

Contact - Deb Agarwal - LBL

Website - <http://www-itg.lbl.gov/Collaboratories/>

[3] Collaboratory Interoperability Framework Project (CIF)

(CIF will provide fundamental communications needs to build on)

Contact - Deb Agarwal - LBL

Website - <http://www-itg.lbl.gov/CIF/>

[4] Advanced Visualization Communications Toolkit

(This will provide network aware components to optimize visualization)

Contact - Deb Agarwal - LBL

Website - <http://www-itg.lbl.gov/~deba/NGI/AdvVizComm.html>

[5] Corridor One Project

(High performance visualization over very high bandwidth networks)

Contact - Rick Stevens - ANL

Website <http://www-fp.mcs.anl.gov/fl/research/Proposals/co.htm>

[6] Toolkit for Collaboratory Development

(Includes Core2000 – a collaboratory research environment based on Habanero from NCSA)

Contact – Jim Meyers – PNL

Website - <http://www.emsl.pnl.gov:2080/docs/collab/>

[7] DOE2000 Electronic Notebook Project

(Electronic notebook projects at PNL, LBL, and ORNL)

Main Website - <http://www.epm.ornl.gov/enote/>

Contacts

Jim Myers - PNL

Sonia Sachs- LBL

Al Geist – ORNL

[8] DOE2000 Collaboratory Research

(Basic collaboratory research projects)

Website - <http://www-unix.mcs.anl.gov/DOE2000/collabs.html>

[9] DOE2000 Collaboratory Pilot Projects

The Diesel Combustion Collaboratory

Website - <http://www-collab.ca.sandia.gov/Diesel/ui/>

The Materials MicroCharacterization Collaboratory

Website - <http://tpm.amc.anl.gov/MMC/>

Environmental Molecular Sciences Collaboratory

Website - <http://www.emsl.pnl.gov:2080/docs/collab/>

Fusion Collaboratory

Website - <http://www.fusionscience.org/collab/REE/>

NSF and Universities

[10] WebFlow

(A visual programming paradigm for Web/Java based coarse grained distributed computing. This is based on Java technology.)

Contacts

Tomasz Haupt - Syracuse University

Wojtek Furmanski - Syracuse University

Website - <http://www.npac.syr.edu/users/haupt/WebFlow/>

[11] SCIRun

(This provides computational steering)

Contact - Christopher R. Johnson - University of Utah

Website - <http://www.cs.utah.edu/sci/publications/scitools96/>

[12] NCSA Collaboration Systems

(Habanero provides state and event synchronization for multiple copies of a software tool. It utilizes Java.)

Contact - Polly Baker - NCSA

Website - <http://havefun.ncsa.uiuc.edu/>

[13] PUNCH

(An Architecture for Web-Enabled Wide-Area Network-Computing)

Contact - Prof Jose Fortes - Purdue Univ

Website - <http://punch.ecn.purdue.edu/>

[14] DISCIPLE

(Distributed System for Collaborative Information Processing and Learning)

Contact - Dr. Ivan Marsic - Rutgers Univ

Website - <http://www.caip.rutgers.edu/multimedia/groupware/>

[15] Space Physics and Aeronomy Research Collaboratory

(an environment for collaborative research in space physics and aeronomy)

Contact - Gary Olson - Univ of Michigan

Website - <http://intel.si.umich.edu/sparc/>

[16] Stanford Interactive Workspaces

(For exploring possibilities for people to work together in technology -rich spaces)

Contact - Terry Winograd - Stanford

Website - <http://graphics.stanford.edu/projects/iwork>

[17] High Performance and Real Time Corba

(Research on improving the throughput performance and reducing latency of Corba)

Contact - Doug Schmidt - UC Irvine (Wash Univ.)

Website - <http://www.cs.wustl.edu/~schmidt/corba-research-performance.html>

NASA

[18] Intelligent Synthesis Environment (ISE) and Collaborative Engineering Environment (CEE)
(NASA's project to create collaborative analysis and design environments)

Contact – W Lundy, NASA Lewis Research Center, for ISE

Website - <http://ise.nasa.gov> Contact - Ed Chow, NASA Jet Propulsion Laboratory, for CEE

Website - <http://ce-server.jpl.nasa.gov/>

[19] Science Desk

(A project to create collaborative research environments with AI support)

Contact - Rich Keller – NASA Ames Research Center

Website - <http://sciencedesk.arc.nasa.gov>

[20] Mars Web Pages

(A website for the collaborative selection of Mars landing sites)

Contact - Glenn Deardorff – NASA Ames Research Center

Website – <http://marsoweb.nas.nasa.gov/landingsites/>

[21] FAST (Flow Analysis Software Toolkit

(A tool for visual analysis of computer simulations of complex physics)

Contact – Tim Sandstrom, NASA Ames Research Center

Website – <http://www.nas.nasa.gov/Software/FAST>

[22] RemoteFAST and FASTexpeditions

(Tools for asynchronous and synchronous collaborative scientific visualization)

Contact – Val Watson – NASA Ames Research Center

Website <http://www.nas.nasa.gov/Software/FAST/FASTexpeditions>

Industry

[23] Intelligent Human-Computer Interaction

(An environment for collaboration based on rooms. Awareness and privacy issues are addressed.)

Contact - Samuel Bayer - Mitre Corp

Website - <http://www.mitre.org/resources/centers/it/g063/hci-index.html>

Commercial CPSE Systems

[24] Tango Interactive

(Based on WebFlow from Syracuse University)

Contact - Marek Podgorny - WebWisdom

Website - <http://www.webwisdom.com/tangointeractive/>

Collaborative Environment Organizations

[25] Computingportals

Home Page - <http://www.computingportals.org/> Survey of projects - <http://www.computingportals.org/projects>

[26] Computer Supported Cooperative Work (CSCW)

Website - <http://www.acm.org/sigchi/cscw2000/index.html>

Reports on Collaborative Environments

[27] Report on Collaborative Virtual Environments 1998

University of Manchester, UK, 17-19th June 1998

Elizabeth Churchill and David Snowden

<http://www.fxpai.xerox.com/ConferencesWorkshops/cve/Report.htm>

[28] Workshop on CPSEs for Scientific Research

San Diego, CA., 29 June – 1 July, 1999

<http://www.emsl.pnl.gov:2080/docs/cpse/workshop/index.html>

Collaborative Environment Standards

[29] International Telecommunication Union's Proposed Standards

Complete listing of proposed standards - <http://www.itu.int/publications/telecom.htm>

Proposed standard for application sharing – <http://www.itu.int/itudoc/itu-t/rec/t/t120.html>

Proposed standard for audiovisual and multimedia systems - <http://www.itu.int/itudoc/itu-t/rec/h/h323.html>

Other references

[30] Kabrisky, Matthew; "A Proposed Model for Visual Information Processing in the Human Brain", University of Illinois Press, 1966